CLAIMS

What is claimed is:

1   1.   A method of identifying a significant phrase in a document, the method

2   comprising:

3   reading a sequence of words from the document;

4   determining a score for each word in the sequence based on the length of each word;

5   comparing the score for each word in the sequence against a threshold score;

6   indicating that the sequence of words is a significant phrase if the number of words in the

7   sequence that have the score greater than the threshold score equals or exceeds a predetermined

8   number;

9   retrieving a sentence from the document, the sentence containing the sequence of words,

10   if the sequence of words is a significant phrase; and

11   searching an abstract of the document to determine whether the sentence is included in

12   the abstract.

1

1   2.   The method of claim 1, wherein reading the sequence of words in the document

2   further comprises reading the sequence of words in the document until a phrase delimiter is

3   detected.

1

1   3.   The method of claim 1, wherein the score for the word is proportional to the

2   number of characters in the word.

1

1   4.   The method of claim 3, wherein the number of characters in the word includes the

2   number of numeric digits in the word.

1

1   5.   The method of claim 1, wherein the score for the word is increased for each

2   capitalized letter in the word.

1

1   6.   The method of claim 1, wherein the score for the word is the length of the word

2   plus the number of capitalized letters in the word.

1

1

1

1

1

2

3

4

1

1

1   7.   The method of claim 1, wherein determining the score for each word in the

2   sequence further includes:

3   determining whether the word exists in a predetermined table; and

4   retrieving the score for the word from the predetermined table.

1   8.   The method of claim 1, wherein the threshold score is based on an average of the

2   scores of words in the sequence of words.

1

1   9.   The method of claim 8, wherein the threshold score is adjusted by a

2   predetermined value to be different from the average of the scores of words in the sequence of

3   words.

1

1   10.   The method according to claim 1, further comprising including the sentence in the

2   abstract, if the sentence is not included in the abstract.

1

1   11.   The method according to claim 1, further comprising discarding the sentence if

2   the sentence is included in the abstract.

1

1   12.   The method according to claim 1, wherein the abstract is language independent.

1

1   13.   A method of identifying a significant phrase in a document, the method

2   comprising:

3   reading a sequence of words from the document;

4   determining a score for each word in the sequence based on the length of each word;

5   comparing the score for each word in the sequence against a threshold score;

6   indicating that the sequence of words is a significant phrase if the number of words in the

7   sequence that have the score greater than the threshold score equals or exceeds a predetermined

8   number;

9   storing the sequence of words and the number of words in the sequence, if the sequence

10 of words is a significant phrase.

1

1   14.   The method according to claim 13, further comprising:

2   retrieving a sentence from the document, the sentence containing the sequence of words,

3   if the sequence of words is a significant phrase; and

4   searching an abstract of the document to determine whether the sentence is included in

5   the abstract.

1

1   15.   The method according to claim 14, further comprising including the sentence in

2   the abstract, if the sentence is not included in the abstract.

1

1   16.   The method according to claim 14, further comprising discarding the sentence if

2   the sentence is included in the abstract.

1

1   17.   The method according to claim 14, wherein the abstract is language independent.

1

1   18.   A computer readable medium containing executable instructions which, when

2   executed in a processing system, cause the system to perform a method for identifying a

3   significant phrase in a document, the method comprising:

4   reading a sequence of words from the document;

5   determining a score for each word in the sequence based on the length of each word;

6   comparing the score for each word in the sequence against a threshold score;

7   indicating that the sequence of words is a significant phrase if the number of words in the

8   sequence that have the score greater than the threshold score equals or exceeds a predetermined

9   number;

10   retrieving a sentence from the document, the sentence containing the sequence of words,

11   if the sequence of words is a significant phrase; and

12   searching an abstract of the document to determine whether the sentence is included in

13   the abstract.

1

1   19.   The computer readable medium according to claim 18, wherein reading the

2   sequence of words in the document further comprises reading the sequence of words in the

3   document until a phrase delimiter is detected.

1

1   20.   The computer readable medium according to claim 18, wherein the score for the

2   word is proportional to the number of characters in the word.

1

1   21.   The computer readable medium according to claim 20, wherein the number of

2   characters in the word includes the number of numeric digits in the word.

1

1   22.   The computer readable medium according to claim 18, wherein the score for the

2   word is increased for each capitalized letter in the word.

23.   The computer readable medium according to claim 18, wherein the score for the

word is the length of the word plus the number of capitalized letters in the word.

24.   The computer readable medium according to claim 18, wherein determining the

score for each word in the sequence further includes:

    determining whether the word exists in a predetermined table; and

4   retrieving the score for the word from the predetermined table.

1

1   25.   The computer readable medium according to claim 18, wherein the threshold

2   score is based on an average of the scores of words in the sequence of words.

1

1   26.   The computer readable medium according to claim 25, wherein the threshold

2   score is adjusted by a predetermined value to be different from the average of the scores of

3   words in the sequence of words.

1

1   27.   The computer readable medium according to claim 18, further comprising

2   including the sentence in the abstract, if the sentence is not included in the abstract.

1

1   28.   The computer readable medium according to claim 18, further comprising

2   discarding the sentence if the sentence is included in the abstract.

1

1   29.   The computer readable medium according to claim 18, wherein the abstract is

2   language independent.

30.   A computer readable medium containing executable instructions which, when

executed in a processing system, cause the system to perform a method for identifying a

significant phrase in a document, the method comprising:

          reading a sequence of words from the document;

          determining a score for each word in the sequence based on the length of each word;

          comparing the score for each word in the sequence against a threshold score;

7   indicating that the sequence of words is a significant phrase if the number of words in the

8   sequence that have the score greater than the threshold score equals or exceeds a predetermined

9   number;

10   storing the sequence of words and the number of words in the sequence, if the sequence

11   of words is a significant phrase.

1

1   31.   The computer readable medium according to claim 30, wherein the method

2   further comprises:

3   retrieving a sentence from the document, the sentence containing the sequence of words,

4   if the sequence of words is a significant phrase; and

5   searching an abstract of the document to determine whether the sentence is included in

6   the abstract.

1

1   32.   The computer readable medium according to claim 31, wherein the method

2   further comprises including the sentence in the abstract, if the sentence is not included in the

3   abstract.

1   33.   The computer readable medium according to claim 31, wherein the method

2   further comprises discarding the sentence if the sentence is included in the abstract.

1   34.   The computer readable medium according to claim 31, wherein the abstract is

2   language independent.

1